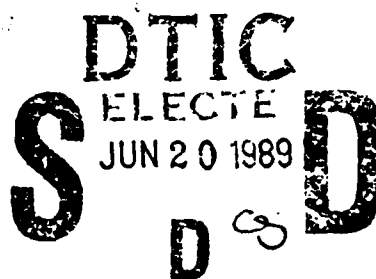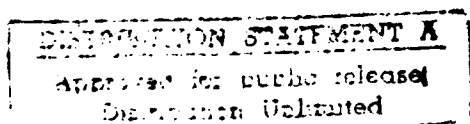AD-A209 840

DTIC
ELECTE
JUN 20 1989
S
D
D

Optimization of Parameters for Semiempirical Methods I. Method

JAMES J. P. STEWART, *Frank J. Seiler Research Laboratory, United States Air Force Academy, Colorado Springs, Colorado 80840*

89   6   19   101

Optimization of Parameters for Semiempirical Methods
I. Method

FJSRL-JR-89-000

James J. P. Stewart

Frank J. Seiler Research Laboratory, United States Air Force Academy, Colorado Springs, Colorado 80840

A new method for obtaining optimized parameters for semiempirical methods has been developed and applied to the modified neglect of diatomic overlap (MNDO) method. The method uses derivatives of calculated values for properties with respect to adjustable parameters to obtain the optimized values of parameters. The large increase in speed is a result of using a simple series expression for calculated values of properties rather than employing full semiempirical calculations. With this optimization procedure, the rate-determining step for parameterizing elements changes from the mechanics of parameterization to the assembling of experimental reference data.

INTRODUCTION

The evolution of molecular orbital (MO) methods has resulted in two main methods, *ab initio* and semiempirical. Of these, *ab initio,* having no need for empirically determined parameters, is the more theoretically pure. Such methods, however, are currently very slow, and routine application at any reasonable degree of accuracy to systems of even a few tens of atoms is still not practical. In contrast, semiempirical methods employ empirically determined parameters and are thus less attractive to the theoretical purist. Unlike *ab initio* methods, the accuracy of any empirical method is limited to the accuracy of the experimental data used in obtaining the parameters. However, semiempirical methods are fast enough and accurate enough for routine application to quite large systems. In consequence, a few semiempirical methods, in particular modified neglect of diatomic overlap (MNDO),[1] have become very popular. To a large degree, this is a result of their computational simplicity as well as their chemically useful accuracy.[2]

The power of semiempirical methods lies not in their theoretical rigor, but in the fact that adjustable parameters within the methods are optimized to reproduce important chemical properties. A comparison of *ab initio* and semiempirical methods has shown that, for heats of formation of systems related to those for which the semiempirical methods were parameterized, the accuracy of semiempirical methods is comparable with that of quite large basis set *ab initio* calculations.[3]

Any semiempirical method is composed of a theoretical framework and a set of parameters. Ideally, these parameters should be fully optimized using an overdetermined reference set of molecules, but hitherto this has not proven possible. In consequence, the quality of semiempirical methods has not been properly tested. It is impossible to partition errors in calculated quantities between deficiencies in the theory and incompleteness of the optimization of the parameters. Only when the parameters for a semiempirical method have been fully optimized can the quality of the theoretical framework be determined. In the limit, when fully optimized parameters are used, the quality of a method depends solely on the nature of the approximations used, and more sophisticated methods will be the more accurate.

This is exemplified by comparison between MNDO and CNDO[4] (complete neglect of differential overlap). MNDO is a more sophisticated method than CNDO, taking into account lone-pair/lone-pair repulsions; therefore, a properly parameterized MNDO model should perform better than an equivalent CNDO model. However, until the parameters within a given method are fully optimized, such a "method" must be regarded as the

CCC 0192-8651/89/020209-12$04.00

combination of the theoretical framework and the current set of parameters. Once the parameters for a method are fully optimized, then reference to the "method" will mean the theoretical framework only.

BACKGROUND TO PARAMETERIZATION

Many elements have been optimized in the MNDO method. For the earlier elements parameterized, a modified form[5] of Bartels'[6] method was used. This general function optimization procedure did not make use of any derivatives. At that time derivatives methods were not practical: "The amount of computation time required would become excessive even for molecules of average size."[1] In consequence, elements were parameterized only with great difficulty. Following the initial parameterization of four elements (hydrogen, carbon, nitrogen, and oxygen), several other elements were parameterized rapidly. In these later parameterizations the values of the parameters for earlier elements were kept constant. A result of this serial parameterization is that only for the set of elements first parameterized could the parameters be fully optimized and internally consistent. Thus the parameters of bromine[7] were optimized to reproduce bromine chemistry while holding the values of all earlier parameters constant. The operation of parameterizing elements is extremely tedious and simultaneous optimization of several elements, other than the initial C—H—N—O set, has not been attempted.

Historically, in order to allow facile comparison of results, the number of sets of MNDO parameters has been kept to an absolute minimum. Only when a readily identifiable and significant deficiency has been identified and corrected has a new set of parameters been published. The set of parameters reported here is in keeping with this philosophy in that they correct various errors which are present when the basic MNDO parameters are used.

Optimization of parameters has proved to be a difficult operation for three reasons. First, there has often been a paucity of accurate experimental data. In certain cases, such as simple gas-phase organic compounds in their ground state, good experimental data exist,[8] but the necessary accurate experimental data for other compounds are frequently absent. Second, optimization of parameters is very time consuming, involving complete geometry optimizations for a large number of molecules at the semiempirical level. Third, the parameters are all highly interdependent. With a nearly optimized set of parameters, any significant change in any parameter *must* be accompanied by a corresponding change in several other parameters. Optimization of the parameters one at a time would be prohibitively slow. This interdependency exists not only within the parameters of a single element, but between parameters of different elements. For example, the optimized value of the orbital exponent for hydrogen is directly affected by the value of the orbital exponent for carbon.

A new and completely general optimization procedure for parameterizing semiempirical methods of the type used in MOPAC[9] is described here. In this method, use is made of the fact that over a sufficiently small domain the calculated values of properties depend linearly on the values of the parameters used. This allows calculated quantities to be approximated by a simple sum of constants times parameters. At intervals, explicit semiempirical calculations which may involve geometry optimization are needed to monitor the validity of this approximation, and to correct accumulated errors. Given a set of experimental data, the method described here can be used to obtain a fully optimized set of parameters relatively rapidly.

THEORY

It is important to know how each parameter is used within the set of approximations used by MNDO and AM1.[10] Thus a brief outline of each parameter will now be given. A more complete description of these methods can be found in Refs. 1 and 10.

For MNDO the calculated heat of formation is given by

$$\Delta H_f = E_{\text{Elect}} + E_{\text{Nuc}} - \sum_A E_{el}^A + \sum_A \Delta H_f^A$$

(1)

where ΔH_f^A is the experimental heat of formation for atom A. E_{el}^A is the calculated en-

ergy of formation of the gaseous atom A from its ion,

$$E_{el}^A = f(P_{ii}, U_{ii}, \langle ii|jj \rangle, \langle ij|ij \rangle), \qquad (2)$$

where P_{ii} is the ground state atomic orbital population of atomic orbital i. The U_{ii} are the one-electron energies of the atomic orbitals of the ion resulting from removal of all valence electrons. These are represented by the parameters U_{ss} and U_{pp}. $\langle ii|jj \rangle$ are the two-electron one-center integrals. In general, there are five of these per atom; these are represented by the parameters: $\langle ss|ss \rangle = G_{ss}$, $\langle ss|pp \rangle = G_{sp}$ $\langle pp|pp \rangle = G_{pp}$, $\langle pp|p'p' \rangle = G_{p2}$, and $\langle sp|sp \rangle = H_{sp}$. Here p and p' are p-type atomic orbitals on a single center but differing in their angular functional form.

E_{Nuc} is the nuclear–nuclear repulsion energy and is given by

$$E_{Nuc} = \sum_{i<j} E_N(i,j) \qquad (3)$$

with

$$\begin{aligned}
E_N(i,j) = \; & Z_A Z_B \langle AA|BB \rangle \\
& \cdot (1 + e^{(-\alpha_A R_{ij})} + e^{(-\alpha_B R_{ij})}) \\
& + (Z_i Z_j / R_{ij}) \\
& \cdot \left(\sum_k a_{kA} e^{-b_{kA}(R_{ij}-c_{kA})^2} \right. \\
& \left. + \sum_k a_{kB} e^{-b_{kB}(R_{ij}-c_{kB})^2} \right), \qquad (4)
\end{aligned}$$

in which atom i is of type A, atom j is of type B, $\langle AA|BB \rangle$ is a two-center integral of type $\langle ss|ss \rangle$, Z_i is the number of valence electrons on atom i, and R_{ij} is the interatomic distance. Only the first term in this expression is present in MNDO. The second term was introduced in the AM1 method in order to correct for excessive long-range repulsions in the core–core repulsion. The quantities α_A, a_{ik}, b_{ik}, and c_{ik} are optimizable parameters.

There are minor variations on this expression when O—H or N—H interactions are involved; these are detailed in Ref. 1.

The total electronic energy is given by

$$E_{Elect} = (1/2)\mathbf{P}(\mathbf{H} + \mathbf{F}), \qquad (5)$$

with $\mathbf{P}$, $\mathbf{H}$, and $\mathbf{F}$ being the density, one-electron, and Fock matrices, respectively. The Fock matrix, over atomic orbitals, is composed of one- and two-electron terms. Diagonal Fock matrix elements are defined by

$$\begin{aligned}
\mathbf{F}_{\mu\mu} = \; & U_{\mu\mu} + \sum_B V_{\mu\mu,B} + \sum_\nu P_{\nu\nu} \\
& \cdot [\langle \mu\mu|\nu\nu \rangle - \langle \mu\nu|\mu\nu \rangle] \\
& + \sum_B \sum_\lambda^B \sum_\sigma^B P\lambda\sigma\langle \mu\mu|\lambda\sigma \rangle, \qquad (6)
\end{aligned}$$

where $V_{\mu\mu,B}$ is the potential experienced by atomic orbital ϕ_μ due to all nuclei in the system other than that of ϕ_μ. Matrix elements connecting orbitals on different atoms are defined by

$$\mathbf{F}_{\mu\nu} = S_{\mu\nu}(\beta_\mu + \beta_\nu)/2 - \sum_{\lambda\sigma} P_{\lambda\sigma}\langle \mu\lambda|\nu\sigma \rangle, \qquad (7)$$

where β_μ and β_ν are atomic parameters of type β_s or β_p. The overlap $S_{\mu\nu}$ is the overlap of two Slater atomic orbitals of type

$$\phi_\mu = N r^{n-1} e^{-\xi r} Y_l^m(\Theta,\phi), \qquad (8)$$

in which ξ is an atomic orbital exponent of type ξ_s or ξ_p. Integrals of the type $\langle \mu\lambda|\nu\sigma \rangle$ are complicated functions of the Slater orbital exponents and the two-electron one-center integrals. The simplest of these two-center terms is the $\langle ss|s's' \rangle$ integral, the repulsion between two electrons each in an s orbital. This integral has the form

$$\langle ss|s's' \rangle = \cfrac{27.21}{\sqrt{\cfrac{1}{R} + \cfrac{1}{4}\left(\cfrac{27.21}{G_{ssA}} + \cfrac{27.21}{G_{ssB}}\right)^2}}. \qquad (9)$$

All other two-center two-electron integrals involve the Slater orbital exponents. A full description of these functions is given in Ref. 11.

In the preceding discussion 12 variables were identified per element within MNDO which could be regarded as parameters. In earlier work, only seven of these were optimized: these are given the symbols U_{ss}, U_{pp}, β_s, β_p, ξ_s, ξ_p, and α. The remaining five, G_{ss}, G_{sp}, G_{pp}, G_{p2}, and H_{sp}, while adjustable parameters in principle, were assigned values determined from atomic spectra.[12] The AM1 method introduced a second set of parameters in the form a gaussian core–core interaction to correct for excessive long-range repulsion in the original MNDO core–core repulsion term. By assigning up to four such gaussians to each atom, as many as 12 more parameters are introduced. Each element in AM1 is described by 10 to 19 parameters.

MNDO-PM3

The theoretical framework for AM1 is similar to that of MNDO, the only difference being the presence in AM1 of radial gaussians which modify the core–core repulsion term. As there have been two earlier parameterizations of the MNDO-type model, in order to facilitate discussion, the resulting method, namely, the combined theoretical framework and optimized parameter set, is best described as a modified neglect of diatomic overlap, parametric method 3, or MNDO-PM3.

In this work, the parameter set will consist of 18 parameters for each element, with the exception of hydrogen, which will have 11 parameters. For convenience, the parameters, $P = P_1, \ldots, P_i, \ldots, P_N$ ($i = 1, N$), are always represented in the following sequence: $U_{ss}, U_{pp}, \beta_s, \beta_p, \xi_s, \xi_p, \alpha, G_{ss}, G_{sp},$ $G_{pp}, G_{p2}, H_{sp}, a_1, b_1, c_1, a_2, b_2, c_2$. For hydrogen the seven parameters involving p orbitals are absent.

REFERENCE FUNCTIONS

As the objective of optimizing parameters is to accurately reproduce experimentally observed values of chemical properties, these quantities must be used in the optimization procedure. This is accomplished by using reference functions which represent experimentally observed phenomena. Reference functions are dimensionless quantities which can be identified with an experimentally observable phenomenon. An example would be the reference function which represents the heat of formation of water. The experimental reference function is obtained by multiplying the observed heat of formation by a weighting factor in (kcal/mol^{-1}). The resulting quantity is, as mentioned above, dimensionless.

Molecular properties are used to define reference functions. The parameters are adjusted to minimize the square of the difference between the calculated and experimental reference function. Within this parameterization method three types of reference functions are used.

Experimental Reference Functions

Most reference functions are derived from experimental data. In a few instances, however, high-level *ab initio* results are used

where experimental data, particularly for geometries, is either absent or of low quality. As with previous optimizations, the sets of properties to be reproduced consist of heats of formation, dipole moments, ionization potentials, and molecular geometries. In addition, heats of reaction and heats of formation for parts of molecules were used.

On occasion, a heat of reaction is known with higher accuracy than the heats of formation of the reactants or products. In these instances, use of heats of formation is unsatisfactory, and the heat of reaction may be used instead.

There are a number of n-alkanes for which the heats of formation are known. The difference in heats of formation of the higher homologs rapidly converges to a constant value: -4.7 kcal/mol per methylene unit. Rather than use all available data on the higher alkanes, a reference function can be constructed from the difference in heats of formation of two adjacent n-alkanes. By assigning an increased weight to this reference function the higher alkanes can be represented. The increased weight will not necessarily result in a reduced error for the n-alkanes used as reference functions. However, the systematic error, that is, the difference in error between the various higher alkanes, will be reduced.

All reference functions used were for properties of gas-phase systems, with the exception of geometric parameters derived from x-ray measurements on solids. In general, compilations of reference data were used; these appeared to have higher internal consistency. Heats of formation used experimentally were relative to the elements in their standard state except for phosphorus, for which white phosphorus was used instead of red. Where possible, experimental ionization potentials (IP) were those obtained by photoionization rather than by electron impact. Where possible, geometries derived from microwave studies were used in preference to x-ray.

As mentioned earlier, reference functions must be made dimensionless in order to be compared. This is accomplished by multiplying the value of each reference function by a constant or weighting factor, w_i, so that the reference data used in the optimization is dimensionless. For heats of formation,

dipoles, and IPs, the default weights were 1.0 kcal/mol^{-1}, 20 D^{-1}, and 10 eV^{-1}, respectively. It is inconvenient to use experimental geometries as reference functions. Instead, the derivative of the heat of formation with respect to geometry for the reference geometry was used. This derivative will be approximately proportional to the amount by which the predicted and observed geometries differ; the proportionality constant being the force constant for that deformation. Only when the predicted and observed geometries are identical will the derivative become zero. The default value of w_i chosen for geometries was 0.7 kcal/mol^{-1} per Å. For certain functions a higher or lower weighting factor was used, depending on their chemical importance. Thus w_i for the heat of formation of benzene was increased to reflect its importance as the prototypical aromatic compound. For similar reasons w_i for the heat of formation of a methylene group in an alkane was also increased.

Calculated Reference Functions

The values of molecular properties calculated using the current set of parameters are also used for constructing reference functions. Ideally, these calculated reference functions would be equal in value to the experimental reference functions.

With the exception of geometric reference functions, all calculated reference functions are evaluated using fully optimized geometries. During the initial stages of the parameter optimization, the geometries changed relatively rapidly, although the total change was, with few exceptions, quite small. The starting geometries generated using MNDO parameters were quite close to both the experimental geometries and to the geometries calculated using the final optimized parameters.

Heat of formation and dipole reference functions, q_i, were derived from the self-consistent field (SCF) heats of formation and dipoles as

$$q_i = w_i \Delta H_{fi}^{calc}$$
$$q_i = w_i D_i^{calc}. \qquad (10)$$

Ionization potential reference functions were simply derived from the molecular orbital energy levels via

$$q_i = -w_i \varepsilon(-n)_i, \qquad (11)$$

where n applies to the ionization potential, and $\varepsilon(-n)$ is the MO energy level $n - 1$ below the highest occupied energy level. In the case of doublet states, a correction was made for the half-electron approximation.

Geometric reference functions were derived from the calculated derivatives of the ΔH_f with respect to the appropriate geometric coordinate using experimental geometries, that is;

$$q_i = -w_i \frac{\delta \Delta H_f}{\delta x_i}, \qquad (12)$$

where x_i is an experimentally determined coordinate, either a bond length, an angle, or a dihedral angle.

Simplified Reference Functions

The form in which the parameters are normally used (as in the equations given above) has serious limitations. The effect of modifying parameters without performing a full MNDO calculation involving SCF calculations and, in certain cases, even geometry optimizations, is impossible to determine. It is desirable, therefore, to represent the MNDO equations in a simpler form, and, where possible, to use this simpler form for all operations involving modification of the parameters. For our purposes, then, the dependency of reference functions which are derived from, for example, heats of formation, will be expressed in terms of the parameters used in the description of the method as a simple summation, that is

$$q_i(\Delta P) = c_i + \sum_j \sum_k a'_{ijk} \Delta P_j^k, \qquad (13)$$

where c_i is the calculated heat of formation, and ΔP_j is the difference in parameter j between the point at which q_i was evaluated, P_j^o, using MNDO and the current value of P_j, that is, $\Delta P_j = (P_j^o - P_j)$.

If only small changes in the values of parameters are to be considered, then this power series can be truncated so that all powers greater than one are ignored:

$$q_i(\Delta P) = c_i + \sum_j a_{ij} \Delta P_j. \qquad (14)$$

DERIVATIVE OF REFERENCE FUNCTION WITH RESPECT TO PARAMETERS

Some derivatives are zero: for a molecule with a given geometry there is no dependency of the ionization potential on the core–core repulsion. Theoretically, these derivatives are finite. In the example given the core–core repulsion affects the geometry which, in turn, affects the IP. However, any significant change in the core–core potential is compensated for by equivalent and opposite changes induced by the other parameters. This state of balance is a consequence of the constraints imposed on the system by the geometry reference functions; the optimized geometries are typically within 2% of the observed values. As a result, such derivatives, although finite, can safely be ignored. Some derivatives can be obtained analytically. For example, the derivative of an ionization potential reference function with respect to the 's' one-center one-electron parameter for element k can be easily obtained from the expression for the MO energy level. For closed-shell systems, Koopmans' theorem[13] states that the ionization potential, I_i, is the negative of the associated molecular orbital energy

$$-I_i = \varepsilon_i = \langle \psi_i | H + F | \psi_i \rangle. \qquad (15)$$

in which the molecular orbitals are normalized LCAOs,

$$\psi_i = \sum_\lambda c_{\lambda i} \phi_\lambda. \qquad (16)$$

For IPs the reference function is equivalent to $-10\varepsilon_i$, therefore

$$\delta q_i / \delta U_{ss}(k) = -10 \sum_\lambda c_{\lambda i}{}^2 \delta(\lambda, s_k), \qquad (17)$$

where $\phi(\lambda, s_k) = 1$ if λ represents an 's' atomic orbital of type k, and zero otherwise.

Only a few derivatives can be calculated using analytical techniques; most derivatives can only be obtained using finite-difference techniques. Table I indicates how each type of derivative is obtained.

RELATIONSHIPS BETWEEN REFERENCE FUNCTIONS

By using dimensionless reference functions and expressing their calculated values as a simple sum, one can determine the similarity or difference of any two reference functions. For both the optimization of parameters and establishing the generality of the method, as wide a range as possible of reference functions must be used.

Each simplified reference function gives rise to a vector which defines a point in parameter space. Although, in principle, exactly n reference functions would be sufficient to define the values of a set of n parameters, in practice many more reference functions than parameters are needed. Inaccuracies in the values of experimental data may bias the parameters. By using a large number of reference functions, errors in any given reference function are minimized. More importantly, reference functions are not independent. That is, two reference functions may have a similar dependency on the parameters. This phenomenon can be quantified using the parameter dependency functions derived above.

For any reference function the parameter dependency can be expressed as a vector of length equal to the number of parameters. To determine the similarity of two reference functions their associated parameter dependency vectors must be normalized. This is done in two steps. First, since the number of times a particular element occurs in a compound varies, the derivatives for each element are individually normalized, thus:

$$q_i(\Delta P) = c_i + \sum_j a_{ij} \Delta P_j, \qquad (18)$$

Table I. Method used for calculation of first derivatives.

Parameter	Heat of formation	Dipole	IP	Geometry
U_{ss}	A	F	F	O
U_{pp}	A	F	F	O
β_s	A	F	F	F
β_p	A	F	F	F
ξ_s	A	F	F	F
ξ_p	A	F	F	F
α	A	O	O	O
G_{ss}	F	F	F	O
G_{sp}	F	F	F	O
G_{pp}	F	F	F	O
G_{p2}	F	F	F	O
H_{sp}	F	F	F	O
a_{ik}	A	O	O	F
b_{ik}	A	O	O	F
c_{ik}	A	O	O	F

Abbreviations: A = analytical derivative; F = finite-difference derivative; O = derivative is zero.

$$\frac{\delta q_i}{\delta P} = \sum_j a_{ij} = \sum_k \sum_{i \in k} a_{ij} , \qquad (19)$$

where k is the index of an element present in the reference function q_i.

Let $c_{ij}' = n_{ik} a_{ij}$ for each element, where n_{ik} is a normalization constant for element k in reference function i. Then

$$\phi_i' = \sum_j c_{ij}' \qquad (20)$$

is a vector representing the reference function in which the coefficients of each element are normalized.

ϕ_i' is a vector of length equal to the number, n_t, of types of elements in the compound. This vector may be normalized thus:

$$\phi_i = \sum_j c_{ij} = n_t^{-1/2} \phi_i' . \qquad (21)$$

Using these vectors the similarity or overlap of any pair of reference functions can be determined. The overlap of two reference functions, $\langle \phi_i | \phi_j \rangle$, is given by

$$\langle \phi_i | \phi_j \rangle = \sum_l c_{il} c_{jl} . \qquad (22)$$

Note that by representing the reference functions by the vectors ϕ_i, all chemical attributes have been removed. This allows a simple comparison to be made between superficially very different forms of data, for example the ionization potential of methanol and the $C-O-C$ angle in dimethyl ether.

Two reference functions which have no elements in common will have an overlap of zero. Conversely, the overlap of two similar compounds, for example the ΔH_f for hexane and heptane, would have an overlap approaching unity. Two reference functions may be described as similar if their overlap is large, greater than about 0.999, and different if their overlap is small, less than about 0.8.

To adequately sample parametric space with a minimum number of reference functions, reference functions which are similar should not be used.

For any given set of reference functions, the degree of independence of parameters can be determined by constructing and diagonalizing the secular matrix of such overlaps. Using m reference functions and n parameters, a secular matrix of size m by m can be constructed. Since there are only n possible independent variables, diagonaliza-

tion of this matrix will yield at least $(m-n)$ exactly zero eigenvalues and at most n positive eigenvalues. These latter n eigenvalues provide an indication of the independency of the parameters. Consider a single-element, n-dimensional, parameter space and n reference functions. If these parameters were fully independent, that is, the overlap of any two reference functions being zero, then all eigenvalues would be identically unity. If the parameters are interdependent, then some roots would be larger than one, while others become less. If any two parameters have exactly the same behavior, then one of the n roots would become exactly zero. For convenience in expressing the interdependency of parameters, the ratio of the total number of reference functions to the smallest eigenvalue will be used.

The principal desideratum of a reference function is that it decrease this ratio. Any reference function which does not contribute significantly to the lowest eigenvector will not affect its value significantly. Consequently, removal of such a function will decrease the number of reference functions by unity, with a concomitant decrease in the interdependency of the parameters.

THE ERROR FUNCTION

With one qualification, a sufficient and necessary condition for an optimized set of parameters is that the error function, S, defined as the sum of the squares of the differences between calculated and experimental values for the reference functions, should be a minimum. That is, given

$$S = \sum_i (q_i^{calc} - q_i^{exp})^2 \qquad (23)$$

where q_i^{calc} and q_i^{exp} are the calculated and experimental values for reference functions, then the parameter set is optimized when S is a minimum.

If only a small number of reference functions are used, then multiple minima are possible. The presence of false minima can be detected by surveying a large number of reference functions. If the parameters are trapped in a local false minimum, then there will be a set of reference functions which will be badly predicted. By including some reference functions from this set, these false

minima may be eliminated. There is, unfortunately, no simple test which indicates when the global rather than local minimum has been reached other than searching for reference functions which are badly predicted.

With this caveat regarding local minima, the task of optimizing parameters is defined. It should be emphasized that no criteria for the values of the parameters other than minimizing S are necessary. The specification of a reference function is sufficiently flexible to allow extra weight to be given to a particular property of a particular compound or reaction. As the stated objective is to predict the values of chemical properties, the values of the parameters as such are of little importance. In particular, any relationship between parameters and quantities of theoretical significance would be fortuitous. This is exemplified by the fact that several low-lying eigenvalues result from diagonalizing the matrix of second derivatives of the reference functions. Small eigenvalues indicate that large changes may be made to the values of the parameters with only a small change resulting in the value of S.

Although the parameters cannot be easily related to quantum-mechanical quantities, in a limited number of cases some physical meaning can be associated with them. The gaussians, which modify the core–core interaction, are an example, Their purpose is to modify the interatomic interaction over the distances within which chemical bonds are important, and secondarily to modify "long-range" interactions up to approximately 3 Å. A large, narrow gaussian centered in this latter region would not be physically meaningful.

Since the reference functions are experimental, that is, they take into account all quantum mechanical operators, and the quantum mechanical framework of MNDO is limited, each parameter which represents a quantum mechanical operator will also represent all its operator equivalents. Thus the carbon 's' orbital exponent, while in principle representing the $2s$ orbital on carbon, will also represent all higher 's' orbitals. Similarly, the gaussian core–core interactions which, in AM1, were intended to correct for an excessive long-range repulsion are also operator equivalents to the Van der Waals attraction or dispersion effects and can therefore behave as if they were dispersion forces

operators. In other words, the correlation effects, which in *ab initio* methods require extensive calculations of the Møller-Plesset type, can, at the MNDO level of approximation, be represented by simple gaussian functions. Conversely, any phenomenon which cannot be represented by any of the operators in MNDO cannot be reproduced. Examples of such phenomena are given in the Discussion section.

OPTIMIZATION OF PARAMETERS

For quadratic functions, minimization is obtained in one step by the operation

$$\mathbf{P}(\text{opt}) = \mathbf{P} - \mathbf{g}/\mathbf{H} \qquad (24)$$

where $\mathbf{H}$ is the Hessian matrix of second derivatives, and $\mathbf{g}$ is the derivative vector $\delta S/\delta \mathbf{P}$.

Although both $\mathbf{g}$ and $\mathbf{H}$ can be readily calculated, the ratio of highest to lowest eigenvalues of $\mathbf{H}$ is very large —when a large number of parameters are involved, this ratio can exceed 10^{12}— and the inverse of $\mathbf{H}$ is numerically very sensitive to small changes in parameters. In part, this is a consequence of the neglect of higher-order terms in the simplified reference function, which preclude the possibility of negative eigenvalues arising from the Hessian of S. This, in turn, leads to completely unrealistic values for $\mathbf{g}/\mathbf{H}$, and the 'optimized' parameters are invariably worse than the initial set.

Several straightforward techniques exist for function optimization: the DFP[14] the Murtagh-Sargent[15] and the BFGS.[16] Each method determines a search direction based on the local gradient and an approximation to the inverse Hessian. In all these methods the accuracy of the inverse Hessian improves as the optimization proceeds.

An initial attempt was made to optimize parameters using the DFP update formula. The time per function evaluation—in this case a full geometry optimization followed by derivative calculation—was prohibitive. The time required for the geometry optimization step alone was approximately 24 hours for the final reference function basis set used, starting from standard bond lengths and angles. Direct function evaluation was clearly impractical. Therefore an alternative

method for determining how to modify parameters had to be developed.

For small changes in parameters, the value of the reference functions, and consequently S, can be calculated using Eq. (14). This allows for rapid evaluation of S at points other than that of the initial function evaluation. In addition, since the local gradient function can be calculated from Eq. (14), not only can a line search be carried out but new search directions can be calculated. In its current form, however, the optimization cannot be allowed to proceed to completion for two reasons: (a) the function S has quadratic dependency on P over only a small domain around the point of the full function evaluation, and (b) in the limit, the numerical instability of the inverse Hessian would result in unrealistic changes in the parameters. The addition to S of a quadratic term of form $c(\Delta p^2) \cdot (P^k - P^{k-1})^2$, where ΔP is the change in P from the initial point, and $P^k - P^{k-1}$ is the change in the values of the parameters between the start of the previous cycle of optimization and the start of the current cycle, moderated the excursion of the P vector. The multiplier c was initially set to an arbitrary positive value and subsequently modified in order to maximize the rate of optimization. Various empirical rules were evolved to determine the value of c on any given cycle. The final set used are as follows:

a. If the final predicted decrease in S and the amount obtained by a full function evaluation were within 25%, the value of c was halved.

b. If the angle between the ΔP vectors on two successive cycles was within 5°, c was halved.

c. If the final decrease in S calculated from a full function evaluation was less than 50% of that predicted, the value of c was doubled.

d. If the angle between the ΔP vectors on two successive cycles was greater than 30°, c was doubled.

After the parameters are optimized, subject to this perturbation, a new explicit function evaluation is done. Because there is *de facto* no error in either S or g at this new point, the origin of the quadratic restraining term used in the previous cycle is moved to the current point. The rapid optimization is

then resumed. Finally, during the optimization some parameters tended to quite unrealistic values. This was a natural consequence of the low eigenvalues for the parameter Hessian matrix. Although, as we have seen, no physical meaning can be attached to any given parameter, it was nevertheless considered appropriate to apply boundary conditions to all parameters. It is likely that parameter sets which contain unrealistic values will eventually be found to be unsuitable for use in the study of certain phenomena, where a more realistic parameter set would be found to be suitable. For each parameter an upper and lower bound was assigned. If the value of a parameter exceeded its allowed bound, the function value within the DFP was increased by 100 times the square of the amount by which the defined bound was exceeded. This sequence was continued until the parameters converged within a preset limit. Using this procedure the sequence for optimizing parameters becomes:

1. Full calculation of S and g.
2. Line search optimization of S using Eq. (14) and a quadratic restraining term.
3. New g evaluated by direct differentiation of Eq. (14) and the quadratic term at the minimum located using step 2.
4. The approximate inverse Hessian updated according to the DFP formula.
5. Steps 2, 3, and 4 repeated until the parameters are converged within a preset limit. The sequence 2–3–4 constitutes one cycle of optimization. Since the first derivative of the energy with respect to optimized geometry is zero, and the IPs and dipoles are relatively insensitive to small geometry changes, molecular geometries were only optimized every few cycles.
6. Unless the change in parameters is acceptably small, the sequence is repeated starting from step 1.

During the initial stages of an optimization, the value of S dropped relatively rapidly. This corresponded to motion in P space along the higher eigenvectors. Once these gross changes were made, the rate of optimization dropped. The average decrease in the value of S on any function evaluation

within the DFP procedure was approximately 0.001%. About 1500 function evaluations were performed during step 5 in the above sequence. Since each function evaluation involved only a double sum, the time required for each of these rapid function evaluations was short. With 400 reference functions and 198 parameters being optimized simultaneously, each line search (which involved typically 5 function evaluations and calculation of derivatives) required 5 s on a Gould MV10000.

SPECIFICATION OF THE MOLECULAR BASIS SET

For each element an initial trial set of molecules and ions was chosen to represent its chemistry. After the parameters were optimized, the full range of compounds were surveyed. Compounds which had large differences between calculated and experimental values for properties were included in the basis set and the parameters reoptimized.

Initially, no organic polynitro compounds were involved. After optimizing the parameters using this deficient basis set, a survey was carried out. This revealed that heats of formation for organic polynitro compounds were very badly predicted. Some representative polynitro compounds were added to the basis set, and the parameters reoptimized. At the end of the optimization, it was found that the S for the original basis set had not risen, but the errors for polynitro organics had dropped considerably. The fact that the original parameter set was optimized and yet a markedly different parameter set could give almost identical results for the original basis set was attributed to the presence of almost zero eigenvalues in the parameter Hessian: motion in any of the directions of the almost zero eigenvectors would not affect the S significantly. Clearly the original large errors in the polynitro compounds were due to the parameter set being insensitive to perturbations in certain directions using the original basis set. On adding polynitro compounds to the basis set this insensitivity was removed, and a new, more general, minimum was reached.

PARAMETER SETS

During the optimization, constraints were imposed on the allowed values of parameters. Even when a large number of molecules are used for defining the values of parameters, the potential still exists for an unrealistic parameter set to be generated. In particular, if the width, parameter 'b,' of the core–core interaction gaussians [Eq. (4)], is unconstrained, then during the initial phase of parameterization these 'b' values could rise to unrealistic values. Inspection of the optimized gaussian invariably revealed that it was centered at a distance corresponding to a specific bond length in a specific compound. When that gaussian was removed, only one compound was affected, and the error in the calculated heat of formation of that compound changed from an initial value of zero to typically a few tens of kcal/mole. To prevent this, the 'b' values were constrained to the range 2.0–6.0 Å.

For iodine, two completely stable sets of parameters were found. No path was found by which one set of parameters could be converted into the other set without passing through a region of very high S. One set corresponded to a large ξ_s and small ξ_p, while the other corresponded to small ξ_s and large ξ_p. Since the two sets were separated by a barrier, and since a large ξ_s would be expected for iodine, the first set of parameters is likely to be the more correct.

No complications were encountered with parameter sets for the other elements. The final values for all parameter sets are presented in Table II.

TECHNICAL DETAILS FOR USING MNDO-PM3 WITH EXISTING PROGRAMS

In order for calculations using MNDO-PM3 to be carried out using earlier copies of MOPAC (Versions 3.1 or higher) or AMPAC, the key words EXTERNAL and AM1 must be used. MNDO-PM3 parameters for an element have two gaussians, therefore it would be necessary to set gaussians 3 and 4 to zero. The parameter names for the parameters given in Table II are, in order, USS, UPP,

Table II. Optimized parameters for MNDO-PM3.

	H	C	N	O	F	Al	Si	P	S	Cl	Br	I
U_{ss}	-13.073321	-47.270320	-49.335672	-86.993002	-110.435303	-24.845404	-26.763483	-40.413096	-49.895371	-100.626747	-116.619311	-96.454037
U_{pp}		-36.266918	-47.509736	-71.879580	-105.685047	-22.264159	-22.813635	-29.593052	-44.392583	-53.614396	-74.227129	-61.091582
Z_s	0.967807	1.565085	2.028094	3.796544	4.708555	1.702888	1.635075	2.017563	1.891185	2.246210	5.348457	7.001013
Z_p		1.842345	2.313728	2.389402	2.491178	1.073629	1.313088	1.504732	1.658972	2.151010	2.127590	2.454354
β_s	-5.626512	-11.910015	-14.062521	-45.202651	-48.405939	-0.594301	-2.862145	-12.615879	-8.827465	-27.528560	-31.171342	-14.494234
β_p		-9.802755	-20.043848	-24.752515	-27.744660	-0.956550	-3.933148	-4.160040	-8.091415	-11.593922	-6.814013	-5.894703
G_{ss}	14.794208	11.200708	11.904787	15.755760	10.496667	5.776737	5.047196	7.801615	8.964667	16.013601	15.943425	13.631943
G_{sp}		10.265027	7.348565	10.621160	16.073689	11.659856	5.949057	5.186949	6.785936	8.048115	16.061680	14.990406
G_{pp}		10.796292	11.754672	13.654016	14.817256	6.347790	6.759367	6.618478	9.968164	7.522215	8.282763	7.288330
G_{p2}		9.042566	10.807277	12.406095	14.418393	6.121077	5.161297	6.062002	7.970247	7.504154	7.816849	5.966407
H_{sp}		2.290980	1.136713	0.593883	0.727763	4.006245	0.919832	1.542809	4.041836	3.481153	0.578869	2.630035
ALP	3.356386	2.707807	2.830545	3.217102	3.358921	1.521703	2.135809	1.940534	2.269706	2.517296	2.511842	1.990185
a_1	1.128750	0.050107	1.501674	-1.131128	-0.012166	-0.473090	-0.390600	-0.611421	-0.399191	-0.171591	0.960458	-0.131481
b_1	5.096282	6.003165	5.901148	6.002477	6.023574	1.915825	6.000054	1.997272	6.000669	6.000802	5.976508	5.206417
c_1	1.537465	1.642214	1.710740	1.607311	1.856859	1.451728	0.632262	0.794624	0.962123	1.087502	2.321654	1.748824
a_2	-1.060329	0.050733	-1.505772	1.137891	-0.002852	-0.154051	0.057259	-0.093935	-0.054899	-0.013458	-0.954916	-0.036897
b_2	6.003788	6.002979	6.004658	5.950512	6.003717	6.005086	6.007183	1.998360	6.001845	1.966618	5.944703	6.010117
c_2	1.570189	0.892488	1.716149	1.598395	2.636158	2.519997	2.019987	1.910677	1.579944	2.292891	2.328142	2.710373

ZSS, ZPP, BETAS, BETAP, GSS, GSP, GPP, GP2, HSP, ALP, FN11, FN21, FN31, FN12, FN22, FN32, and, in order to eliminate the third and fourth gaussian of AM1, FN13, FN14. The end of the oxygen EXTERNAL parameter set would thus be

FN22	0	5.950512
FN32	0	1.598395
FN13	0	0.000000
FN14	0	0.000000

To verify that the parameters have been correctly entered, calculations on representative systems should be done.

CONCLUSION

The new method for optimizing parameters for semiempirical quantum-mechanical methods is several orders of magnitude faster than those previously employed. Highly optimized values for parameters for 12 elements have been obtained which are suitable for use in the study of chemical systems. Although there is no unambiguous way of determining if the current parameter set corresponds to the global minimum, extensive checking of the applicability of the parameters to a wide range of compounds (following article) does not suggest that any significantly lower minimum exists.

The presence of several almost zero eigenvalues in the parameter Hessian matrix indicates that the parameter set is not unambiguously defined, and that caution must be exercised in applying these parameters to the study of new systems. A welcome corollary of this is that if any deficiencies are identified, and that the reference functions representing these deficiencies can increase the value of any of these almost zero eigenvalues, indicating that the position in parameter space of the associated eigenvector

has become defined, then the parameter set can likely be reoptimized so as to remove these deficiencies without compromising the applicability of the parameters to known systems.

References

1. M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, **99**, 4899 (1977).
2. M. J. S. Dewar, *J. Phys. Chem.*, **89**, 2145 (1985).
3. M. J. S. Dewar and D. M. Storch, *J. Am. Chem. Soc.*, **107**, 3898 (1985).
4. J. A. Pople and G. A. Segal, *J. Chem. Phys.*, **43**, S136 (1965); 289 (1966).
5. P. K. Weiner, Ph.D. dissertation, University of Texas at Austin, Austin, Texas, 1975.
6. R. H. Bartels, University of Texas Center for Numerical Analysis, Report CNA-44, Austin, Texas, 1972.
7. M. J. S. Dewar and E. F. Healy, *J. Comput. Chem.*, **4**, 542 (1983).
8. See, for example, J. D. Cox and G. Pilcher, *Thermochemistry of Organic and Organometallic Compounds*, Academic Press, New York, (1970).
9. J. J. P. Stewart, MOPAC, A Semi-Empirical Molecular Orbital Program, *QCPE*, **455** (1983).
10. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.*, **107**, 3902 (1985).
11. M. J. S. Dewar and W. Thiel, *Theoret. Chim. Acta (Berl.)*, **46**, 89 (1977).
12. L. Oleari, L. DiSipio, and G. DeMichelis, *Mol. Phys.*, **10**, 97 (1966).
13. T. Koopmans, *Physica (Utrecht)*, **1**, 104 (1934).
14. R. Fletcher and M. J. D. Powell, *Comput. J.*, **6**, 163 (1963); W. C. Davidon, *ibid.*, **10**, 406 (1968).
15. B. A. Murtagh and R. W. H. Sargent, *Comput. J.*, **13**, 185 (1970).
16. C. G. Broyden, *J. Inst. Math. App.*, **6**, 222 (1970); R. Fletcher, *Comp. J.*, **13**, 317 (1970); D. Goldfarb, *Math. Comput.*, **24**, 23 (1970); D. F. Shanno, *Math. Comput.*, **24**, 647 (1970).